

A Survey on the Memory Mechanism of Large Language Model-based Agents

Year: 2024 | Citations: 274 | Authors: Zeyu Zhang, Quanyu Dai, Xiaohe Bo

Abstract

Large language model (LLM)-based agents have recently attracted much attention from the research and industry communities. Compared with original LLMs, LLM-based agents are featured in their self-evolving capability, which is the basis for solving real-world problems that need long-term and complex agent-environment interactions. The key component to support agent-environment interactions is the memory of the agents. While previous studies have proposed many promising memory mechanisms, they are scattered in different papers, and there lacks a systematical review to summarize and compare these works from a holistic perspective, failing to abstract common and effective designing patterns for inspiring future studies. To bridge this gap, in this article, we propose a comprehensive survey on the memory mechanism of LLM-based agents. In specific, we first discuss “what is” and “why do we need” the memory in LLM-based agents. Then, we systematically review previous studies on how to design and evaluate the memory module. In addition, we also present many agent applications, where the memory module plays an important role. At last, we analyze the limitations of existing work and show important future directions. To keep up with the latest advances in this field, we create a repository at https://github.com/nuster1128/LLM_Agent_Memory_Survey.